**Preprocessing**

<table>
<thead>
<tr>
<th>Grayscale Image</th>
<th>CLAHE</th>
<th>Otsu Threshold</th>
</tr>
</thead>
</table>

**Motivation**

Metastases in lymph nodes are one of the most important prognostic factors in breast cancer. This is because the lymph nodes are the first place the cancer is likely to spread. The current manual diagnostic procedure is time consuming and prone to error.

We present multiple models and feature selection algorithms and compare their results at a difficult multi-class classification task.

**Data**

- Data from the CAMELYON challenge 2017
- 500 images from 100 patients with ground truths for 4 classes
- Giga-pixel slide images in a multi-level format

**Color Buckets**

- Histogram of pixel colors
- Metastases tend to be darker

**Histogram of Gradients**

- Histogram of pixel-wise gradients in 8x8 blocks

**Feature Extraction Methods**

**SURF with Bag of Words**
- Extract features with SURF
- Cluster features (kNN) to get bag of words
- Encode each image using the bag

**Selective Search**
- Propose bounding boxes
- Isolate smaller regions of interest

**Models: SVM**

Fits 4 planes used for binary ‘one-vs-all’ classification among all the classes
- Polynomial Kernel
  \[ K(x,y) = (x^T + c)^d \]
- Hinge loss
  \[ 1 = \max(0, 1-yx) \]

**Models: Ensemble Learning**

Uses multiple ‘weak learners’ which cast weighted votes during classification.

New learners focus on misclassifications
- AdaBoost: Decision trees as weak learners
- Subspace: kNN as weak learners
- RUSBoost: Decision trees as weak learners

**Models: Residual Network**

A convolutional NN with a skip connection
- Each example is comprised of stacked bounding boxes from selective search
- Forward Prop
  \[ z^i = W[i]^T \cdot a^{i-1} + b^{i-1} + Wx \]
  \[ a^i = g(z^i) \]

**Results**

Training on 400 examples. Test on 100 examples.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Feature Extraction</th>
<th>Train/Test Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (polynomial)</td>
<td>Color Buckets</td>
<td>62/63***</td>
</tr>
<tr>
<td>SVM (polynomial)</td>
<td>Bag of Words</td>
<td>55/24</td>
</tr>
<tr>
<td>RUSBoost</td>
<td>HOG</td>
<td>86/24</td>
</tr>
<tr>
<td>Residual Network</td>
<td>Selective Search</td>
<td>80/46</td>
</tr>
</tbody>
</table>

**Discussion**

- Imbalance of classes and size of the images poses a unique problem
- Feature selection made more important
- Generalization appears to be a problem

**Future Work**

- Experiment with ‘one-vs-one’ classification for SVMs to alleviate class imbalance
- Train two classifiers, one binary and one to discriminate types of metastases
- Experiment with better feature selection/region proposal algorithms
  - R-CNN

**References**